Check for updates

A wavelength-scale black phosphorus spectrometer

Shaofan Yuan [1], Doron Naveh [2], Kenji Watanabe [3], Takashi Taniguchi [4] and Fengnian Xia [1] ✉

On-chip spectrometers with compact footprints are being extensively investigated owing to their promising future in critical applications such as sensing, surveillance and spectral imaging. Most existing miniaturized spectrometers use large arrays of photodetection elements to capture different spectral components of incident light, from which its spectrum is reconstructed. Here, we demonstrate a mid-infrared spectrometer in the 2–9 μm spectral range, utilizing a single tunable black phosphorus photodetector with an active area footprint of only 9×16 μm², along with a unique spectral learning procedure. Such a single-detector spectrometer has a compact size at the scale of the operational wavelength. Leveraging the wavelength and bias-dependent responsivity matrix learned from the spectra of a tunable blackbody source, we reconstruct unknown spectra from their corresponding photoresponse vectors. Enabled by the strong Stark effect and the tunable light–matter interactions in black phosphorus, our single-detector spectrometer shows remarkable potential in the reconstruction of the spectra of both monochromatic and broadband light. Furthermore, its ultracompact structure that is free from bulky interferometers and gratings, together with its electrically reconfigurable nature, may open up pathways towards on-chip mid-infrared spectroscopy and spectral imaging.

Spectrometers are cornerstone instruments in scientific research across various disciplines and in technological developments for many industries. Traditional spectrometers are usually bulky and consist of mechanically movable parts, such as optical gratings in visible and near-infrared spectrometers[1] or Michelson interferometers in Fourier transform infrared (FTIR) spectrometers[2]. Miniaturized spectrometers free from mechanical moving parts are being extensively investigated due to their low cost, compactness and potential for in-field applications[3–10]. On-chip FTIR spectrometers based on evanescent standing waves[8] and phase shifters[9] in the near-infrared region have previously been demonstrated on a silicon platform leveraging computational reconstruction methods and usually have an operational wavelength range of twenty to around one hundred nanometres. Compact spectrometers with wider operational wavelength ranges have also been reported by leveraging separate photodetectors and additional optical component arrays with wavelength-dependent transmission or reflection responses, such as arrays of quantum dots[3] and photonic crystals[6] in the visible wavelength range and arrays of metasurfaces[4,10] in the mid-infrared wavelength range. Recently, detector-only spectrometers in the visible range have been reported[5,7]. These on-chip spectrometers consist of arrays of photodetectors with varying optical responsivities to capture the different spectral components of incoming light. These detector-only spectrometers[5,7] have smaller footprints compared with those made from separate photodetection and spectral-selective elements[3,4,6,10]. However, the resolution of this type of spectrometer is still limited by the number of photodetectors in the array. Moreover, their on-chip footprint is considerably larger than its operational wavelength since many photodetectors are used in this scheme. In the previously demonstrated compact spectrometers that cover the visible[3,5–7], near-infrared[8,9] and mid-infrared[4,10] regions, the spectra of the incident light are calculated by leveraging the measured photoresponses and reconstruction algorithms.

In this work, we report a new scheme for mid-infrared spectroscopy using a single tunable black phosphorus (BP) photodetector, whose footprint is comparable to its operational wavelength. BP has high carrier mobilities, a moderate bandgap (0.3 eV) in its bulk form and a high optical absorption, which make it a promising candidate for mid-infrared photoconductors[11–13]. Our new scheme for mid-infrared spectroscopy is enabled by the strong photoresponse and Stark effect in BP devices[14–18], which is made possible by incorporating a mid-infrared transparent graphene top gate[19,20], trap-free hexagonal boron nitride (hBN) encapsulation[21] and a BP thin film of appropriate thickness. The spectrometer demonstrated in this work is almost maximally miniaturized considering the diffraction of light, and thus future spectral imaging sensors will not have substantially improved spatial resolution even with further reduction in the pixel size[22]. Different from other on-chip spectrometers based on detector arrays, which are fixed after fabrication[5,7], our device can be electrically reconfigured after fabrication. This single-detector scheme also potentially has a higher sensitivity compared with detector array schemes, since in this scheme the incident light is focused on a single element instead of being split to the array of detectors[5,7]. Moreover, the intrinsic spectral properties of the BP spectrometer have been 'learned' from a blackbody source with tunable temperature. The entire scheme does not rely on the traditional optical components such as interferometers, gratings or tunable lasers. Thus, our demonstration provides a potentially simplified and inexpensive route for mid-infrared spectroscopy and spectral imaging.

As a single device, a compact spectrometer with moderate resolution can make the spectral information more accessible for sensing applications. As discussed later in this work, from the moderate-resolution absorption spectra, the BP spectrometer can clearly capture the characteristic absorption peak of carbon dioxide (CO_2) at around 4.3 μm. It also successfully determines the
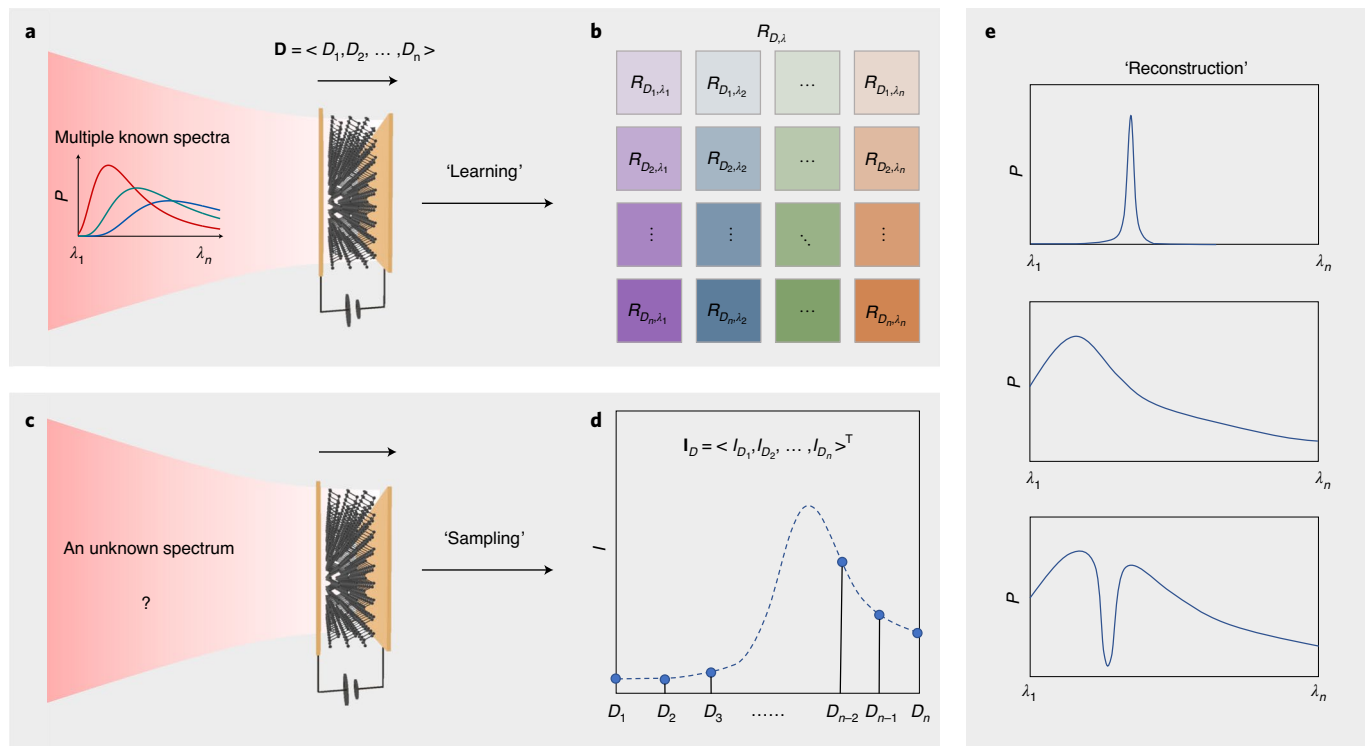
Fig. 1 | The operational principle of the single-detector BP spectrometer consists of three steps: learning, sampling and reconstruction. a, Schematic of the learning process. The photoresponse of BP (the dark grey layered material) can be tuned by the electric displacement (*D*) between two parallel metallic plates. **b**, The responsivity matrix $R_{D,\lambda}$ of the BP spectrometer, which depends on the biasing displacement field (*D*) and the incident light wavelength (λ), is 'learned' from the photoresponses to a group of known spectra (illustrated by the red, green and blue curves in **a** under different biasing displacement fields). **c**, Schematic of the sampling process. For the incident light with an unknown spectrum, the photocurrent in the BP detector is sampled as a function of the electric displacement *D*. **d**, The sampling process generates the photocurrent vector **I$_D$**. **e**, Schematics of three reconstructed spectra. From top to bottom, the spectra represent a narrowband emission, a broadband emission and an absorption feature of certain gas molecules. Reconstruction of the unknown spectra leverages the $R_{D,\lambda}$ and **I$_D$** as discussed in **a**–**d**.

spectral lineshape of blackbody radiation with moderate resolution. Moreover, the reported single-device scheme makes the realization of an array of such spectrometers possible, which may enable the future spectral imaging in the mid-infrared. Previously, additional grey-scale thermal images in the mid-infrared acquired on automobiles, drones or satellites have been utilized for identifying pedestrians, vehicles and hazardous gas[23–25]. Compared with these greyscale thermal images, a compact spectral imager can offer higher detection ability since the spectral information can be continuously measured, albeit with moderate resolution. Thus, the compact spectral imager with moderate resolution can be leveraged for applications such as in the identification of objects and threats. In addition, it can also be useful in remote sensing[26].

Operational scheme

The operational scheme of our single-detector BP spectrometer is illustrated in Fig. 1. It consists of three steps: (1) learning, (2) sampling and (3) reconstruction. The optical properties of the BP thin film (the layered material in Fig. 1a) can be tuned by an external biasing displacement field (*D*), which is controlled by voltages applied to the pair of metallic plates in Fig. 1a. The photoresponsivity (*R*) of this BP photodetector is a function of both the incident light wavelength (λ) and the external biasing displacement field *D*.

In the first stage, the spectrometer responsivity *R* as a function of both displacement *D* and wavelength λ is learned. The continuous responsivity function *R*(*D*,λ) can be discretized into a matrix $R_{D,\lambda}$ in Fig. 1b. As will be detailed later, at each displacement field D_i, by measuring the photoresponses to multiple known incidence

spectra (the inset of Fig. 1a), the responsivity row vector for D_i, $(R_{D_i,\lambda_1}, R_{D_i,\lambda_2}, \ldots, R_{D_i,\lambda_n})$, in the matrix $R_{D,\lambda}$ can be inferred through a learning process. Completing the learning process for all *n* displacement fields generates the full *n* by *n* matrix in Fig. 1b.

In the next sampling step as shown in Fig. 1c,d, the photocurrent response to the incident light with an unknown spectrum will be measured at *n* different displacement fields (D_1 to D_n), generating the response vector **I$_D$**. Based on the responsivity matrix $R_{D,\lambda}$ generated by the learning process (Fig. 1b) and the response vector **I$_D$** (Fig. 1d) measured in the sampling step, the spectrum of the incident light can be reconstructed. From top to bottom, Fig. 1e illustrates three representative spectra reconstructed in this work, which are a narrowband emission, a broadband emission and an absorption feature of certain gas molecules, respectively, from top to bottom.

Characterization of transport and optical properties

Figure 2a shows a schematic of the BP spectrometer based on the dual-gate transistor configuration (see the inset of Fig. 2c for a top-view optical image). The active region consists of an approximately 13-nm-thick BP thin film sandwiched between the top and bottom hBN layers (see Supplementary Fig. 1 for the thickness measurements). The source–drain electrodes are composed of chromium/gold. The top hBN layer is capped by monolayer graphene, which is isolated from the source–drain electrodes and functions as the top-gate metal. The entire device is fabricated on a silicon dioxide-covered silicon substrate, which is used as the bottom gate (see Methods for details of device fabrication). The top and bottom gates can independently control the biasing
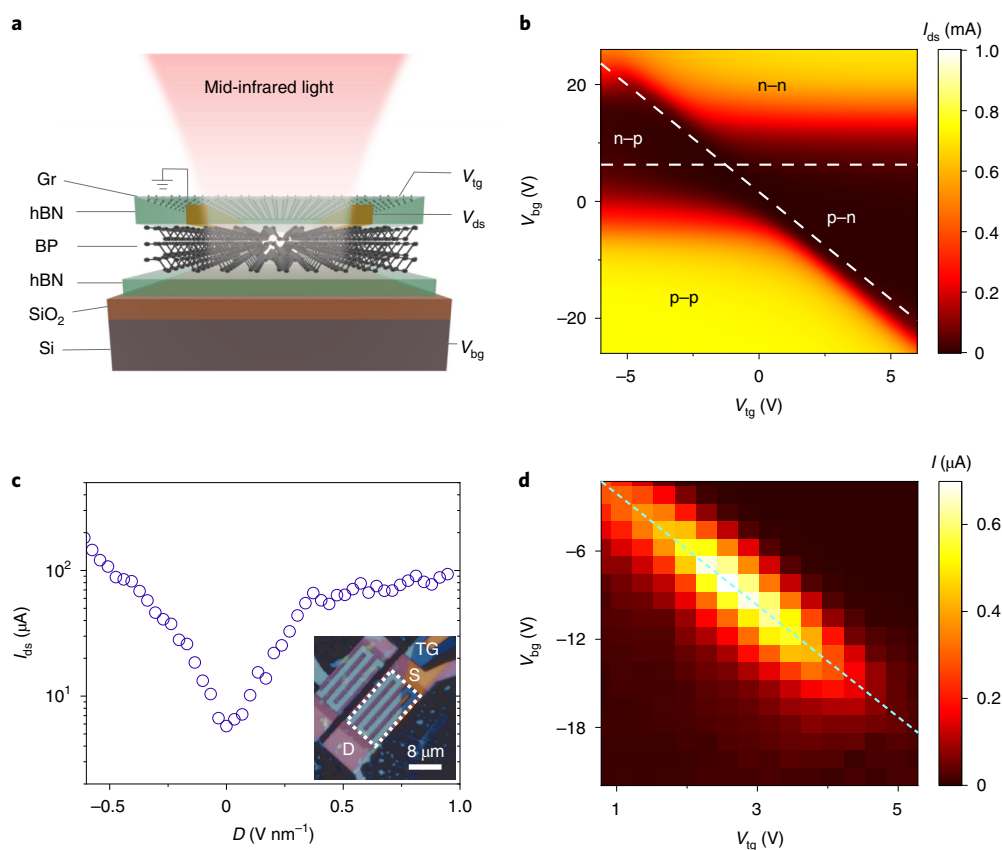
Fig. 2 | Characterization of the tunable BP spectrometer. a, Schematic of the BP spectrometer. It consists of a hBN/BP/hBN heterostructure capped by monolayer of graphene on a silicon dioxide-covered silicon substrate. Graphene and the silicon substrate are used as the top and back gates, respectively. The source–drain electrodes are composed of chromium/gold. Gr, graphene. **b**, The source–drain current (I_{ds}) as a function of the top-gate (V_{tg}) and back-gate (V_{bg}) biases. The I_{ds} mapping is divided by the charge-neutrality line of the BP channel (slanted) and the horizontal line into four areas based on the polarities of the BP in the channel and under the metal contacts. Along the horizontal line, the BP under the source and drain contacts is tuned to intrinsic by the bottom gate regardless of the top-gate bias (V_{tg}) due to the screening of top gating fields by the source and drain electrodes. **c**, I_{ds} as a function of the displacement field D along the charge-neutrality line. Inset: optical image of two BP spectrometers, with the active area ($9 \times 16\,\mu m^2$) marked with white dashed lines for one of the devices; D, S and TG represent the drain, source and top-gate electrodes, respectively. **d**, The measured photocurrent as a function of V_{tg} and V_{bg}. Maximum photocurrent is measured along the charge-neutrality (slanted) line.

displacement fields in the top- and bottom-gate dielectric layers, respectively. Encapsulation by hBN prevents the oxidation of BP, and the trap-state-free surfaces of hBN also minimize trap-induced photocurrents in the photoresponses[21,27]. As a result, our BP spectrometer operates in intrinsic photoconduction mode and has negligible hysteresis. Compared with the previously reported tunable BP photodetector[17], there are two main differences. First, in this device we used mid-infrared-transparent monolayer graphene as the top gate. In contrast to the previously used ultrathin metal top gate[17], the graphene monolayer is very transparent to mid-infrared light with >99% transmission due to Pauli blocking[19,20]. Second, in this device, the BP is around 13 nm thick, which is much thicker than that used in ref. [17]. The performance of the spectrometer critically depends on the BP thickness. Light absorption in thinner BP is weaker and it also takes a larger field to tune its bandgap[17,18]. However, if the BP thickness is too large (close to or above 20 nm), the free-carrier screening effect becomes substantial and the bandgap tuning is again suppressed[18,28]. As a result, a thickness in the range of 10–15 nm is preferred for this demonstration[17,18,28].

Figure 2b plots the source–drain current (I_{ds}) as a function of top-gate (V_{tg}) and bottom-gate (V_{bg}) biases. In this measurement, the source was grounded and the drain was biased at $V_{ds} = 0.5$ V. These transport (and later photocurrent) measurements were conducted at 80 K (see Methods for details of transport and photocurrent measurements). Since liquid nitrogen is widely available, many infrared photodetection elements in commercial spectrometers operate at a temperature of around 77 K (refs. [29,30]), thus an operating temperature of 80 K is acceptable for spectroscopy applications. We can divide Fig. 2b into four regions using two dashed lines based on the polarities of the BP in the transistor channel and under the source and drain electrodes, which is similar to the transport mapping in a previous dual-gate graphene device[31]. In Fig. 2b, the slanted line represents the charge-neutrality condition of the BP channel: along this line, the BP channel is intrinsic since the carriers induced by the top- and bottom-gate biases have opposite polarities and compensate the doping effect, and a conductance minimum is observed. Along the horizontal line, the BP underneath the source and drain contacts is tuned to be intrinsic by the bottom gate regardless of the top-gate bias (V_{tg}), due to the screening of top gating fields by the source and drain electrodes.

Figure 2c shows the source–drain current as a function of the biasing displacement field along the charge-neutrality line in Fig. 2b, where $D = \varepsilon_{tBN}(V_{tg} - V_{t0})/d_{tBN}$. Here, ε_{tBN} (3.1) and d_{tBN} (21 nm) are the permittivity and thickness of top hBN, respectively (see Supplementary Fig. 1 for thickness measurements). The parameter V_{t0} (−0.7 V) is the offset top-gate voltage, which accounts for the

unintentional doping of BP. The source–drain current (I_{ds}) rises as the biasing displacement field increases, which implies the reduction of the BP bandgap[28]. However, accurate determination of the bandgap based solely on these transport results is not feasible due to the unknown contact resistance. As will be discussed later in this work, we determine the spectral response and the bandgap tuning of BP under bias in the next section. In Fig. 2d, we plot the gate-dependent photocurrent under the light excitation from a 1,073 K blackbody light source. The maximum photoresponse is observed along the charge-neutrality line of the BP since the lifetime of photocarriers is the longest when the carrier density is minimized[32]. The photocurrent first increases when the biasing displacement field rises, since under bias the BP absorption edge is extended to longer wavelength range. However, when $V_{tg} > 2.6$ V (or $D > 0.48$ V nm^{-1}), the photocurrent starts to decline. This phenomenon could be due to two reasons. First, although bias reduces the bandgap and extends the cutoff wavelength, it could weaken the oscillator strength and overall less incident light is absorbed by the BP[18]. Second, a reduction in the bandgap leads to a higher intrinsic carrier concentration, which can result in a reduced photocarrier lifetime. In this work, all photocurrent measurements for spectroscopy were performed along the charge-neutrality line. At $V_{bg} = 6$ V, the BP under the source and drain electrodes is intrinsic (there is residue doping such that the charge-neutrality condition does not occur at zero back-gate bias) and the contact resistance is large. The photoresponse is relatively weak in the region where V_{bg} is close to 6 V, because the light excitation only modifies the channel resistance but not the contact resistance. As a result, in Fig. 2d, we scan the photoresponse for V_{bg} from −1 to −21 V, where BP under the source and drain metal is doped by the back-gate biasing fields and the photoresponse is strong.

Spectral response learning process

Traditionally, the spectral responses in the mid-infrared can be characterized using FTIR spectrometers or tunable lasers. Here we use a blackbody source with tunable temperatures to train a responsivity model and acquire the spectral responsivity matrix. In this process, the learning data are the photocurrents ($\mathbf{I}_T$) collected at displacements D excited by a blackbody light source at temperature T. At a given displacement D_i, the photocurrent (I) is expected to depend on the blackbody source temperature T and the spectral responsivity $R(\lambda)$ as $I = f(T, R(\lambda))$, where f is an unknown non-linear function. A regression analysis that captures this non-linear relationship between I and T will allow us to extract $R(\lambda)$. However, establishing an accurate non-linear regression process is challenging. Here, we map the function f to the space of the incident power density $P(T, \lambda)$ and spectral responsivity $R(\lambda)$. The incident power density of the blackbody source depends on the wavelength λ and the temperature T, and it can be calculated based on Planck's Law[33]. The blackbody spectra used in this work deviate from those of an ideal blackbody by less than 1%, and the calibrated emissivity results are traceable to the National Institute of Standards and Technology[34].

In the non-saturated region, the photocurrent $I(T)$ is an integral of the product of the incident spectral power density and the responsivity over the entire wavelength range of interests from λ_1 to λ_n, $I(T) = \int_{\lambda_1}^{\lambda_n} P(T, \lambda) R(\lambda) \, d\lambda$, where $T = T_1, T_2, ... T_n$. The incident spectral power densities from a blackbody source at three typical temperatures, T_i, T_j and T_k, are shown schematically in Fig. 3a. Also shown in Fig. 3a is the unknown schematic responsivity $R(\lambda)$ to be determined at the given displacement D_i in this learning process. For a given temperature T_i, we can measure the photocurrent $I(T_i) = \int_{\lambda_1}^{\lambda_n} P(T_i, \lambda) R(\lambda) \, d\lambda$ collected by the spectrometer, which is the integral of the blue shaded area in Fig. 3b. A fitting to similar

photocurrent measurements at the blackbody temperatures T_i (the blue shaded area), T_j (the green shaded area) and T_k (the red shaded area), and all other temperatures generates the schematic $I(T)$ curve (black line) shown in the inset of Fig. 3b. For blackbody temperatures from T_1 to T_n, we have n integral equations and can decompose them into a matrix equation by discretization

$$(R_{\lambda_1}, R_{\lambda_2}, ..., R_{\lambda_c}) \begin{pmatrix} P_{T_1,\lambda_1} & P_{T_2,\lambda_1} & \cdots & P_{T_n,\lambda_1} \\ P_{T_1,\lambda_2} & P_{T_2,\lambda_2} & \cdots & P_{T_n,\lambda_2} \\ \vdots & \vdots & \ddots & \vdots \\ P_{T_1,\lambda_c} & P_{T_2,\lambda_c} & \cdots & P_{T_n,\lambda_c} \end{pmatrix} = (I_{T_1}, I_{T_2}, ..., I_{T_n})$$

(1)

or $\mathbf{R} \times P_{T,\lambda} = \mathbf{I}_T$ in a more compact form. Details of the discretization process are presented in the Methods section. This is the pivotal equation in the learning process and gives us the responsivity vector $\mathbf{R}_{D_i}$ at displacement field D_i as shown in Fig. 3c. Note that here the responsivity vector ends at λ_c, which is determined by the displacement D_i. To solve the matrix equation (1) and make the solution robust against noise and errors, we used an adaptive Tikhonov regularization method[5,35,36] and introduced the cutoff wavelength of the photoconductors λ_c to stabilize the solution against different types of noise. To test the feasibility of this learning process for the responsivity, we characterized commercial lead selenide and lead sulfide photodetectors using this method and compare the results with those measured using a commercial FTIR spectrometer. Both results are presented in Supplementary Fig. 2 and a good agreement is observed, which indicates the validity of the learning process.

We applied the learning process to generate the spectral responsivity vectors at 41 different displacements fields. The spectral responsivity matrix, $R_{D,\lambda}$, was then reconstructed using these 41 spectral response vectors. Figure 3d shows the photocurrent responses of the BP spectrometer under 41 different biasing displacement fields when it is illuminated by the blackbody source at different temperatures. Figure 3e plots the spectral responsivity matrix as a function of D and λ, which is reconstructed from Fig. 3d after linear averaging. It is clear that the cutoff wavelength of the BP spectrometer is extended from ~4 μm in the minimum biasing condition to around 9 μm at a biasing displacement field of 0.7 V nm^{-1}, thus enabling the broadband mid-infrared spectroscopy measurements.

Spectroscopy demonstrations

With the known spectral responsivity matrix, $R_{D,\lambda}$, we perform the spectrum measurements using our single-detector spectrometer. To reconstruct the unknown spectra, we first measure the photocurrent as a function of the displacement field, as illustrated in the sampling process in Fig. 1c,d. At a given displacement field D_i, the photocurrent can be expressed in an integral $I(D_i) = \int_{\lambda_1}^{\lambda_n} R(D_i, \lambda) \times P(\lambda) d\lambda$, where $P(\lambda)$ is the unknown spectrum to be determined as illustrated schematically by the blue dashed line in Fig. 4a. Here we use a narrowband emission spectrum as an example. Three schematic spectral response curves at three different displacement fields (D_i, D_j and D_k) are plotted schematically in Fig. 4a. For each displacement field, the device shows different photocurrents (red, blue and green crosses in Fig. 4b) determined by the responsivity curves and the incident spectrum. By fitting the photocurrent data (including the red, blue and green crosses in Fig. 4b) as a function of D, the unknown spectrum can be reconstructed as shown in the right panel of Fig. 4b. In this process, similar to the learning process, the continuous integral equations at these displacements were discretized and grouped into a matrix equation,
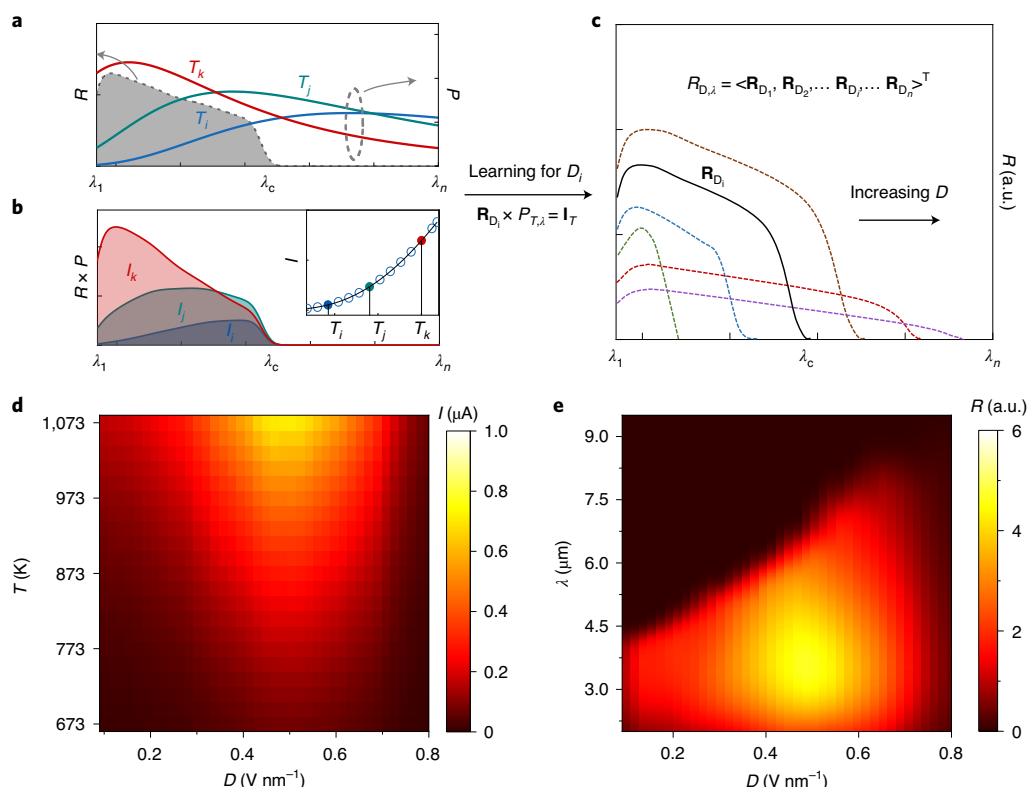
Fig. 3 | Spectral responsivity matrix learning. a, Power spectral densities (*P*) of a blackbody source at three different temperatures (T_i, T_j and T_k, right axis) and the schematic unknown responsivity curve of BP (*R*) to be learned at a given biasing displacement field D_i (the dashed grey curve, left axis). **b**, The photoresponse spectrum (*P × R*) excited by the blackbody source at different temperatures, T_i (blue), T_j (green) and T_k (red). Integration of the photoresponse spectrum leads to the photocurrent. By fitting the photocurrent as a function of the temperature as shown in the inset, the photoresponse spectrum of BP at certain displacement fields (D_i) can be determined. **c**, The photoresponsivity spectrum for a certain D_i (the solid black curve) is reconstructed from the photocurrents and incident light spectra in **a** and **b**. Using similar methods, the photoresponsivity spectra at other electric displacements (dashed curves) can also be reconstructed. The spectral responsivity matrix $R_{D,\lambda}$ can be expressed as a series of spectral responsivity vectors, $\mathbf{R}_{D_i}$. Here $\mathbf{R}_{D_i}$ is a discrete form of the solid back curve in **c**. **d**, The photocurrent acquired at different electric displacements (*D*) and blackbody temperatures (*T*). **e**, The inferred photoresponsivity of BP as a function of *D* and wavelength (λ). $R_{D,\lambda}$ is constructed directly from the data in **d**.

$$\begin{pmatrix} R_{D_1,\lambda_1} & R_{D_1,\lambda_2} & \cdots & R_{D_1,\lambda_n} \\ R_{D_2,\lambda_1} & R_{D_2,\lambda_2} & \cdots & R_{D_2,\lambda_n} \\ \vdots & \vdots & \ddots & \vdots \\ R_{D_n,\lambda_1} & R_{D_n,\lambda_2} & \cdots & R_{D_n,\lambda_n} \end{pmatrix} \begin{pmatrix} P_{\lambda_1} \\ P_{\lambda_2} \\ \vdots \\ P_{\lambda_n} \end{pmatrix} = \begin{pmatrix} I_{D_1} \\ I_{D_2} \\ \vdots \\ I_{D_n} \end{pmatrix}, \qquad (2)$$

or $R_{D,\lambda} \times \mathbf{P} = \mathbf{I}_D$ in a compact form. Similarly, we used regularizations in the regression to avoid overfitting and divergent results in this adaptive reconstruction process (see Methods for details).

We first measured the spectrum of a mid-infrared laser at around 5 μm. Figure 4c shows the photocurrent measured by the spectrometer as a function of displacement *D* from around 0.10 to 0.83 V nm⁻¹. We applied a least absolute shrinkage and selection operator ('Lasso') regression to reconstruct the spectrum, which is designed for sparse signals[37]. The green curve in Fig. 4d shows the reconstructed spectrum, and the dark grey solid line is the spectrum measured using the commercial FTIR spectrometer. The full-width at half-maximum (FWHM) of the reconstructed spectrum is around 500 nm, which is limited by the Nyquist–Shannon sampling theorem[38]. Namely, in the measured wavelength range from 1 to 9.5 μm, the optimal resolution based on the 41 sampling points is 420 nm, which is very close to the measured FWHM. By comparison, the intrinsic FWHM of this laser spectrum is around 15 nm, which is much narrower than the measured result using our

spectrometer, as determined by the commercial FTIR spectrometer. The resolution in our spectrometer can be improved by increasing the sampling photocurrent points and/or by choosing a smaller displacement field scanning range as shown below.

The cyan and red curves in Fig. 4d show the reconstructed laser spectrum with 41 and 81 photocurrent sampling points, respectively, within a displacement scanning range of 0.19–0.56 V nm⁻¹. Such a narrowed scanning range corresponds to the scanning wavelength range of 4–7 μm. With the reduced scanning range and 81 photocurrent sampling points, a FWHM of 90 nm is achieved (red curve), corresponding to a FWHM energy of 4 meV. In this case, the resolution in energy already shows an advantage if compared with the recently demonstrated detector-only spectrometers[5,7]. In the reconstruction process, we use the linearly interpolated data in the responsivity matrix when needed, since the original spectral responsivity matrix has a dimension of 41. Further improvement in the resolution is possible by enlarging the dimension of the responsivity matrix and sampled photocurrent vector. Compared with previously demonstrated detector-only spectrometers, in which the number of detector elements is fixed after fabrication[5,7], we can reconfigure our displacement field tuning range and increase the number of sampling points to improve the resolution in the selected spectral range.

We further demonstrate its capability of reconstructing the spectrum of the broadband emission. Figure 4e shows the displacement-dependent photocurrent excited by the blackbody
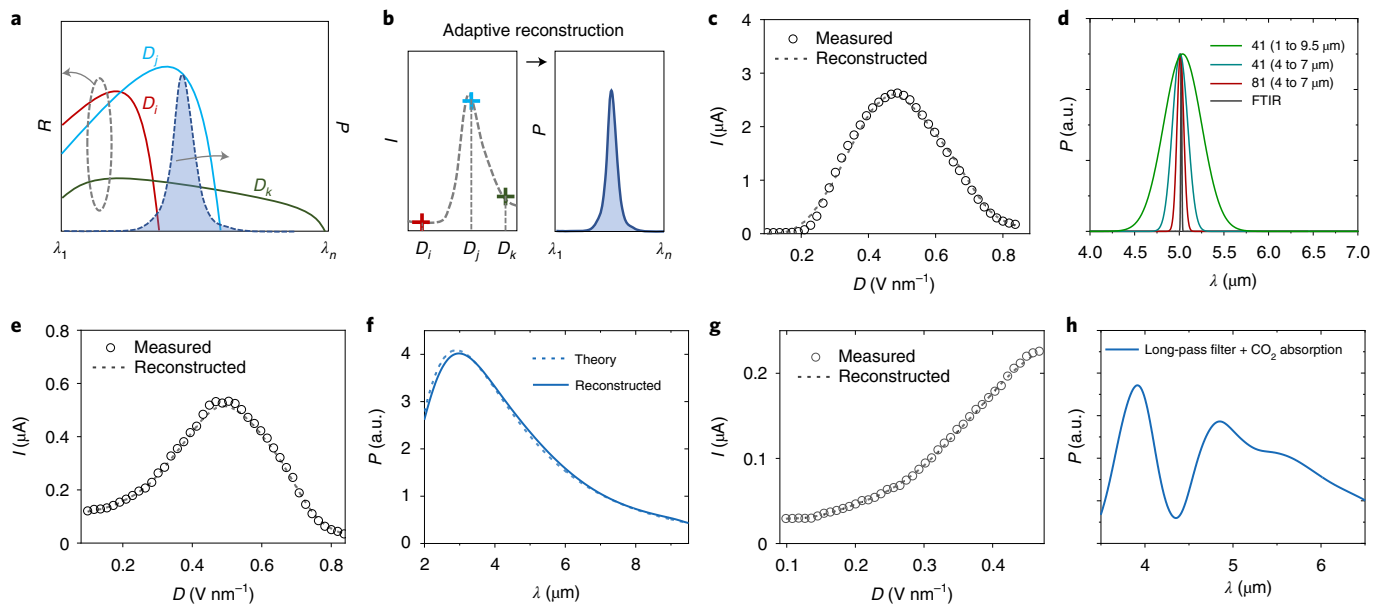
Fig. 4 | Spectroscopy demonstrations. a, Photoresponsivity spectra at three typical electric displacements, D_i (red), D_j (blue) and D_k (green), and a schematic of an unknown incident light spectrum (blue dashed line, right axis). **b**, By fitting the measured photocurrents as a function of D (left panel), including the data measured at D_i, D_j and D_k, the unknown spectrum can be reconstructed (right panel). **c**, The photocurrent excited by a 5 μm infrared laser is measured as a function of D, from $D = 0.10$ to $0.83\ V\ nm^{-1}$ (open circles, 41 points). The dashed line is its corresponding fitting. **d**, The reconstructed spectra (green, cyan and red curves) and the reference spectrum (dark grey) measured using a commercial FTIR spectrometer. The reconstruction wavelength range for the green curve is 1–9.5 μm with 41 points, and it is 4–7 μm for the cyan and red curves with 41 and 81 points, respectively. As we reduce the reconstruction spectral range and increase the number of sampling points, the resolution is improved. **e**, The photocurrent excited by a 1,000 K blackbody source is measured as a function of D (open circles), and the dashed line is its corresponding fitting in the reconstruction. **f**, The reconstructed spectrum for the measured blackbody source (solid curve) and its corresponding theoretical spectrum as given by Planck's Law (dashed curve). **g**, The photocurrent measured for $D = 0.10$–$0.47\ V\ nm^{-1}$ (open circles) when the chamber is filled with CO_2 and with a long-pass filter cutoff at 3.7 μm in the pathway. The dashed line is the corresponding fitting curve in the reconstruction process. **h**, The reconstructed spectrum from **g**. It captures the absorption by CO_2 (the dip at around 4.3 μm) and the extinction edge of the long-pass filter at around 3.7 μm.

source at 1,000 K. The spectrum of the blackbody source at this temperature (1,000 K) was not utilized in the learning process. As a result, we can use it to test the performance of our spectrometer. By leveraging Tikhonov regularization regression[5,35,36], the blue solid curve in Fig. 4f is reconstructed, and it agrees well with the theoretical curve (blue dashed line) calculated using Planck's Law[33]. In this measurement, the displacement field is from 0.1 to 0.83 V nm^{-1} and the number of sampling points is 41. Finally, we filled the optical path in our measurement system with CO_2 and inserted a long-pass filter with a cutoff wavelength at 3.7 μm, then we repeated the spectrum measurement. Figure 4g illustrates the measured displacement-dependent photocurrent for $D = 0.1$–0.47 V nm^{-1} with 41 sampling points. The reconstructed spectrum based on the measurements in Fig. 4g is shown in Fig. 4h. Although the resolution is limited by the dimension of the responsivity matrix and the sampled photocurrent vector, the measured spectrum captures the strong absorption around 4.3 μm by the infrared-active vibrational modes of CO_2 (ref. [39]) and the cutoff wavelength of the filter at around 3.7 μm. We further compare the characteristics of our single-detector spectrometer with some typical compact spectrometers in terms of the operational mechanism, footprint and resolution in Supplementary Table 1. As illustrated in this table, various compact spectrometers have been demonstrated in the visible[3,5–7], near-infrared[8,9] and mid-infrared[4,10] ranges leveraging computational methods, but their footprints are still much larger than the operational wavelengths. By comparison, our demonstration represents a mid-infrared spectrometer with a footprint at the scale of the wavelength. Moreover, the reconfigurability nature of the BP spectrometer can enable dense sampling for a higher resolution within

a desirable spectral range, as demonstrated above in the measurement of the laser spectrum. Based on the reconfigurable nature of the device, more advanced sampling strategies, such as random sampling for compressed sensing[40,41], can potentially be leveraged to achieve a resolution beyond the Nyquist–Shannon limit[38,42].

Finally, we comment on the possible limiting factors on the spectrometer resolution. The first practical factor limiting the sampling points n could be the numerical property of the responsivity matrix. When the photoresponse difference between two nearby sampling points is diminishingly small, the nearby rows in the responsivity matrix $R_{D,\lambda}$ become highly correlated. This could make the matrix ill-conditioned, leading to difficulty in the determination of its inverse. The second factor that could limit the resolution is the size of the training dataset used for determining the responsivity matrix. As a two-dimensional matrix, the number of elements increases quadratically with n. Third, the computational power may limit the maximum number of sampling points (n) in the future. As the number of sampling points (n) increases, the runtime of the algorithm will increase substantially since it grows on the order of the cube of n (ref. [43]). These limiting factors might be alleviated if a deep-learning method is used. Deep learning can be utilized to make predictions without being explicitly programmed to solve the matrix equations in this case[44]. In the training process, parameters in the neural network are adjusted to map inputs to outputs and minimize the difference between the outputs and references. Then, in the reconstruction, it will be able to predict or reconstruct new spectra based on the neural network. As a result, it may not be necessary to deal with the matrix equations directly. Moreover, incremental learning based on neural networks can expand the training

dataset as we do more measurements[45]. In this case, the resolution could be improved when we do more measurements. Finally, once properly trained, the deep-learning system may be able to reconstruct the spectra in real time since there will be no iteration in the reconstruction step, which could greatly reduce the required computation power[46].

Summary

Here we report a mid-infrared spectroscopy scheme based on a single, tunable BP detector. The spectroscopy function has been realized within an active area footprint of merely $9 \times 16\,\mu m^2$, which is comparable to its operational wavelength. The entire measurement scheme does not rely on advanced optical components such as interferometers, movable gratings or tunable infrared lasers. Thus, our demonstration represents a major step towards the extreme miniaturization of spectrometers and may enable on-chip, affordable mid-infrared spectroscopy and spectral imaging.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41566-021-00787-x.

References

1. Hollas, J. M. *High Resolution Spectroscopy* (Butterworth-Heinemann, 1982).
2. Griffiths, P. R. Fourier-transform infrared spectrometry. *Science* **222**, 297–302 (1983).
3. Bao, J. & Bawendi, M. G. A colloidal quantum dot spectrometer. *Nature* **523**, 67–69 (2015).
4. Tittl, A. et al. Imaging-based molecular barcoding with pixelated dielectric metasurfaces. *Science* **360**, 1105–1109 (2018).
5. Yang, Z. Y. et al. Single-nanowire spectrometers. *Science* **365**, 1017–1020 (2019).
6. Wang, Z. et al. Single-shot on-chip spectral sensors based on photonic crystal slabs. *Nat. Commun.* **10**, 1020 (2019).
7. Meng, J. J., Cadusch, J. J. & Crozier, K. B. Detector-only spectrometer based on structurally colored silicon nanowires and a reconstruction algorithm. *Nano Lett.* **20**, 320–328 (2020).
8. Le Coarer, E. et al. Wavelength-scale stationary-wave integrated Fourier-transform spectrometry. *Nat. Photon.* **1**, 473–478 (2007).
9. Kita, D. M. et al. High-performance and scalable on-chip digital Fourier transform spectroscopy. *Nat. Commun.* **9**, 4405 (2018).
10. Shrestha, V. R. et al. Mid- to long-wave infrared computational spectroscopy with a graphene metasurface modulator. *Sci. Rep.* **10**, 5377 (2020).
11. Liu, H. et al. Phosphorene: an unexplored 2D semiconductor with a high hole mobility. *ACS Nano* **8**, 4033–4041 (2014).
12. Li, L. K. et al. Black phosphorus field-effect transistors. *Nat. Nanotechnol.* **9**, 372–377 (2014).
13. Xia, F., Wang, H. & Jia, Y. C. Rediscovering black phosphorus as an anisotropic layered material for optoelectronics and electronics. *Nat. Commun.* **5**, 4458 (2014).
14. Kim, J. et al. Observation of tunable band gap and anisotropic Dirac semimetal state in black phosphorus. *Science* **349**, 723–726 (2015).
15. Liu, Y. P. et al. Gate-tunable giant Stark effect in few-layer black phosphorus. *Nano Lett.* **17**, 1970–1977 (2017).
16. Whitney, W. S. et al. Field effect optoelectronic modulation of quantum-confined carriers in black phosphorus. *Nano Lett.* **17**, 78–84 (2017).
17. Chen, X. L. et al. Widely tunable black phosphorus mid-infrared photodetector. *Nat. Commun.* **8**, 1672 (2017).
18. Chen, C. et al. Widely tunable mid-infrared light emission in thin-film black phosphorus. *Sci. Adv.* **6**, eaay6134 (2020).
19. Mak, K. F., Ju, L., Wang, F. & Heinz, T. F. Optical spectroscopy of graphene: from the far infrared to the ultraviolet. *Solid State Commun.* **152**, 1341–1349 (2012).
20. Yan, H. et al. Infrared spectroscopy of wafer-scale graphene. *ACS Nano* **5**, 9854–9860 (2011).
21. Dean, C. R. et al. Boron nitride substrates for high-quality graphene electronics. *Nat. Nanotechnol.* **5**, 722–726 (2010).
22. Rogalski, A., Martyniuk, P. & Kopytko, M. Challenges of small-pixel infrared detectors: a review. *Rep. Prog. Phys.* **79**, 4 (2016).
23. Sandsten, J., Weibring, P., Edner, H. & Svanberg, S. Real-time gas-correlation imaging employing thermal background radiation. *Opt. Express* **6**, 92–103 (2000).
24. Gaszczak, A., Breckon, T. P. & Han, J. Real-time people and vehicle detection from UAV imagery. *Proc. SPIE* **7878**, 78780B (2011).
25. Hwang, S., Park, J., Kim, N., Choi, Y. & Kweon, I. S. Multispectral pedestrian detection: benchmark dataset and baseline. In *Proc. Conf. on Computer Vision and Pattern Recognition*, 1037–1045 (IEEE, 2015).
26. Underwood, E. C., Ustin, S. L. & Ramirez, C. M. A comparison of spatial and spectral image resolution for mapping invasive plants in coastal California. *Environ. Manag.* **39**, 63–83 (2007).
27. Guo, Q. et al. Black phosphorus mid-infrared photodetectors with high gain. *Nano Lett.* **16**, 4648–4655 (2016).
28. Deng, B. et al. Efficient electrical control of thin-film black phosphorus bandgap. *Nat. Commun.* **8**, 14474 (2017).
29. *FT-IR Spectrometer Detectors* (Newport, 2021); https://www.newport.com/f/detectors-for-oriel-mir8035-ft-spectrometers
30. *VERTEX Series: Advanced Research FT-IR Spectrometers* (Bruker, 2018); https://www.bruker.com/fileadmin/user_upload/8-PDF-Docs/OpticalSpectrospcopy/FT-IR/VERTEX/Brochures/VERTEXseries_Brochure_EN.pdf
31. Gabor, N. M. et al. Hot carrier-assisted intrinsic photoresponse in graphene. *Science* **334**, 648–652 (2011).
32. Selberherr, S. *Analysis and Simulation of Semiconductor Devices* (Springer, 1984).
33. Planck, M. & Masius, M. *The Theory of Heat Radiation* (P. Blakiston's Son & Co., 1914).
34. *SR-2 High Temperature Cavity Blackbody* (CI Systems, 2019); https://www.ci-systems.com/Files/Source%20-%20SR-2%20%20Cavity%20BB.pdf
35. Tikhonov, A. N., Goncharsky, A. V., Stepanov, V. V. & Yagola, A. G. *Numerical Methods for the Solution of Ill-Posed Problems* (Kluwer Academic, 1995).
36. Hansen, P. C. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion* (Society for Industrial and Applied Mathematics, 2005).
37. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* **58**, 267–288 (1996).
38. Proakis, J. G. & Manolakis, D. G. *Digital Signal Processing: Principles, Algorithms, and Applications* 3rd edn (Prentice Hall, 1996).
39. *Carbon Dioxide Infrared Spectrum* (NIST Chemistry WebBook, 2018); https://webbook.nist.gov/cgi/cbook.cgi?ID=C124389&Units=SI&Type=IR-SPEC&Index=1#IR-SPEC
40. Candès, J. E. & Wakin, B. M. An introduction to compressive sampling. *IEEE Signal Process. Mag.* **25**, 21–30 (2008).
41. Eldar, Y. C. & Kutyniok, G. *Compressed Sensing: Theory and Applications* (Cambridge Univ. Press, 2012).
42. Candès, J. E., Romberg, K. J. & Tao, T. Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.* **59**, 1207–1223 (2006).
43. Arora, S. & Barak, B. *Computational Complexity: A Modern Approach* (Cambridge Univ. Press, 2009).
44. Samuel, A. L. Some studies in machine learning using the game of checkers. *IBM J. Res. Dev.* **3**, 210–229 (1959).
45. Polikar, R., Upda, L., Upda, S. S. & Honavar, V. Learn++: an incremental learning algorithm for supervised neural networks. *IEEE Trans. Syst. Man Cybern. C* **31**, 497–508 (2001).
46. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, 2016).

Methods

Device fabrication. The construction of hBN/BP/hBN heterostructures utilizes previously reported polymer-free dry transfer methods[47,48]. A 13-nm-thick BP flake was first mechanically exfoliated and encapsulated by two hBN flakes (bBN and tBN-2) in an argon-filled glovebox. Then the heterostructure was transferred on to a 90 nm silicon dioxide-covered silicon substrate. The top hBN was then etched partially to expose the BP for contact metal deposition. A chromium/gold layer (3/27 nm, respectively) was thermally evaporated to form the source and drain electrodes and metal pads for probing. Then the third hBN flake (tBN-1) and a chemical-vapour-deposited graphene were subsequently transferred on to the device and the graphene was patterned by dry etch to isolate it from the source and drain electrodes.

Transport and photocurrent characterization. The device was loaded in a low-temperature chamber (model HFS600E-PB4 from Linkam Scientific Instruments) for all the transport and photocurrent measurements at 80 K. The dual-gate transport was measured using Keithley 2612B and Keithley 2500 source meters. To measure the photocurrent in BP, the incident light was first chopped by a mechanical chopper at a modulation frequency, then photocurrents at the modulation frequency were collected using a pre-amplifier (Femto DLPCA-200) and lock-in amplifier (Model SR830). In the measurement of the photocurrent excited by the blackbody light source, the low-temperature chamber, the blackbody source (Model SR200 by CI Systems) and all of the optical path were contained in a nitrogen-purged sealed box, which was kept at a positive pressure compared with the ambient environment. In the measurement of the photocurrent excited by the mid-infrared laser, the output light from a quantum cascade laser (5 μm) was coupled to an infrared microscope (Hyperion 2000) and focused on the device.

Discretization of integral equations. A discretization approach reported in ref. [5] is adopted here. The responsivity, $R(\lambda)$, of the spectrometer is a continuous function of the wavelength λ. Since the spectrometer shows little response to light beyond its cutoff wavelength (λ_c), $R(\lambda)$ is close to zero when $\lambda > \lambda_c$. To reconstruct the continuous $R(\lambda)$ using finite numbers of photocurrent measurements $I(T)$, $R(\lambda)$ can be approximated by a linear combination of n Gaussian basis functions, $R(\lambda) \approx \sum_{i=1}^n R_{\lambda_i}\phi_i(\lambda_i)$. Here, $\phi_i(\lambda_i)$ is a Gaussian function with a centre wavelength of λ_i and a standard deviation of $\sigma = (\lambda_{max}-\lambda_{min})/n$, where λ_{max} and λ_{min} represent the upper and lower boundaries of the spectral range, respectively, and R_{λ_i} is the coefficient of term i in the linear combination[5]. At a blackbody temperature T, the excited photocurrent in BP is the integral

$$I(T) = \int_{\lambda_{min}}^{\lambda_{max}} P(T,\lambda) R(\lambda)\, d\lambda$$
$$\approx \sum_{i=1}^n R_{\lambda_i}\int_{\lambda_{min}}^{\lambda_{max}} P(T,\lambda)\phi_i(\lambda_i)d\lambda,$$

where $P(T,\lambda)$ is the temperature- and wavelength-dependent power density of the blackbody source. For simplicity, if we define $P_{T,\lambda_i} = \int_{\lambda_{min}}^{\lambda_{max}} P(T,\lambda)\phi_i(\lambda_i)d\lambda$, the integral turns into a linear combination, $I(T) = \sum_{i=1}^n R_{\lambda_i}P_{T,\lambda_i}$. To train the responsivity model in the learning process, we measured the photocurrent $I(T)$ at different blackbody temperatures from T_1 to T_n and generated the matrix equation in equation (1) in the main text. By solving these equations, $R(\lambda)$ can be reconstructed based on the linear combination of R_{λ_i}. For the highest blackbody temperature $T = 1,073$ K, the incident power at wavelengths below 1 μm is less than 0.1% of the overall incident power. Therefore, we set $\lambda_{min} = 1$ μm and neglect the incident light at shorter wavelengths. The cutoff wavelength λ_c obtained from the cross-validation described in the main text is within 9.5 μm and the BP spectrometer shows little photoresponse at longer wavelengths at all displacement fields. Thus, λ_{max} is set to be 9.5 μm in the regression. In our spectrometer, 9 μm is considered to be the longest wavelength at which the spectrometer can confidently function, since its responsivity from 9 to 9.5 μm is rather weak. The incident power at wavelengths of 1–2 μm accounts for 9% of the total power. This relatively small power results in less accuracy in the spectral responsivity learning process at wavelengths of 1–2 μm. Although these wavelengths are considered in the learning and reconstruction process, the results are not expected to be as accurate as other wavelengths. As a result, 2–9 μm is considered as the operational wavelength range of this spectrometer.

In the reconstruction process, the photocurrent is a function of the electric displacement D, $I(D) = \int_{\lambda_{min}}^{\lambda_{max}} R(D,\lambda) P(\lambda)\, d\lambda$. In this process, the incident power density $P(\lambda)$ is approximated by $P(\lambda) \approx \sum_{i=1}^n P_{\lambda_i}\phi_i(\lambda_i)$. Considering the photocurrent at different electric displacements, the matrix equation in equation (2) in the main text was obtained through a similar discretization process.

Learning and reconstruction processes. Finding the best solution for equation (1) is to find a responsivity vector **R** to minimize the squared error, $\|\mathbf{R}\times P_{T,\lambda}-\mathbf{I}_T\|^2$, between the measurements ($\mathbf{I}_T$) and calculated results, $\mathbf{R}\times P_{T,\lambda}$. The direct answer given by the inverse of matrix $P_{T,\lambda}$ times $\mathbf{I}_T$ usually leads to a divergent result since it amplifies the noise in $P_{T,\lambda}$ (refs. [5,36]). To stabilize the solution and make it robust against noise, we used a regularization method[5,35–37]. In this process, the goal is to identify **R** to minimize the cost function, cost $= \|\mathbf{R}\times P_{T,\lambda}-\mathbf{I}_T\|^2 + \alpha\times w(\mathbf{R})$.

Here, $\alpha\times w(\mathbf{R})$ is the penalty term in the cost function, where $w(\mathbf{R})=\|\mathbf{R}\|^2$ in the Tikhonov regularization[35] or $\|\mathbf{R}\|$ in the Lasso regularization[37].

In the learning process, there were two adaptive parameters, the cutoff wavelength λ_c and the above-mentioned penalty term coefficient α, to avoid overfitting and to make the learning process robust in the presence of measurement noise[36]. Rather than manually choosing these parameters, we used cross-validations to adaptively choose α and λ_c for a specific dataset[49]. The general idea is to find suitable α, λ_c and the corresponding $\mathbf{R} = \langle R_{\lambda_1}, R_{\lambda_2}, ..., R_{\lambda_c}\rangle$ to minimize mean squared errors (MSEs) of the regression in a cross-validation. In the cross-validation, the dataset was divided into a training set and a test set. The model was first trained using the training set and then we calculated the MSEs as a function of these two parameters α and λ_c using the test set, and finally selected the parameters that gave the least MSE. We introduced the cutoff wavelength λ_c because the spectrometer shows negligible response beyond it. An inappropriate λ_c will result in underfitting or overfitting and a large MSE. The α term was also chosen to keep the model from overfitting or underfitting the photocurrent. Intuitively, if α is too large the fitting will be poor. On the other hand, a very small α may lead to better fitting to the training data, but it can also cause the failure of generalization to the test data[36]. Thus, adaptively choosing these parameters is necessary for the training process. Similarly, a penalty coefficient α is also included in the reconstruction process. Tikhonov regularization is chosen for broadband spectra[35] and Lasso regularization is used for narrowband spectra[37]. The power density vector is initialized with the learning data in the reconstruction of the broadband spectra.

Data availability
The data that support the plots within this paper are available from the corresponding author upon reasonable request.

Code availability
The code and algorithm in this paper are available from the corresponding author upon reasonable request.

References
47. Wang, L. et al. One-dimensional electrical contact to a two-dimensional material. *Science* **342**, 614–617 (2013).
48. Chen, X. L. et al. High-quality sandwiched black phosphorus heterostructure and its quantum oscillations. *Nat. Commun.* **6**, 7315 (2015).
49. Golub, G. H., Heath, M. & Wahba, G. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* **21**, 215–223 (1979).

Author contributions
S.Y., D.N. and F.X. conceived the idea and initiated the project. S.Y. performed the experiments, collected all the data and developed the analysing computer programs. K.W. and T.T. synthesized the hBN crystals. All the authors discussed the results and commented on the manuscript.

Competing interests
S.Y., D.N. and F.X. are applying for an international patent based on the spectroscopy scheme and results presented in this work. The remaining authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41566-021-00787-x.

Correspondence and requests for materials should be addressed to F.X.

Peer review information *Nature Photonics* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.